ORIGINAL PAPER

Fitting optimum order of Markov chain models for daily rainfall occurrences in Peninsular Malaysia

**Sayang Mohd Deni · Abdul Aziz Jemain ·
Kamarulzaman Ibrahim**

Abstract The analysis of the daily rainfall occurrence behavior is becoming more important, particularly in water-related sectors. Many studies have identified a more comprehensive pattern of the daily rainfall behavior based on the Markov chain models. One of the aims in fitting the Markov chain models of various orders to the daily rainfall occurrence is to determine the optimum order. In this study, the optimum order of the Markov chain models for a 5-day sequence will be examined in each of the 18 rainfall stations in Peninsular Malaysia, which have been selected based on the availability of the data, using the Akaike's (AIC) and Bayesian information criteria (BIC). The identification of the most appropriate order in describing the distribution of the wet (dry) spells for each of the rainfall stations is obtained using the Kolmogorov-Smirnov goodness-of-fit test. It is found that the optimum order varies according to the levels of threshold used (e.g., either 0.1 or 10.0 mm), the locations of the region and the types of monsoon seasons. At most stations, the Markov chain models of a higher order are found to be optimum for rainfall occurrence during the northeast monsoon season for both levels of threshold. However, it is generally found that regardless of the monsoon seasons, the first-order model is optimum for the northwestern and eastern regions of the peninsula when the level of thresholds of 10.0 mm is considered. The analysis indicates that the first order of the Markov chain model is found to be most appropriate for describing the distribution of wet spells, whereas the higher-order models are found to be adequate for the dry spells in most of the rainfall stations for both threshold levels and monsoon seasons.

S. M. Deni (✉)
Center of Statistical Studies, Faculty of Information Technology
and Quantitative Science, Universiti Teknologi MARA (UiTM),
40450 Shah Alam, Selangor, Malaysia
e-mail: sayan929@salam.uitm.edu.my

A. A. Jemain · K. Ibrahim
School of Mathematical Sciences,
Faculty of Science and Technology,
Universiti Kebangsaan Malaysia (UKM),
43600 Bangi, Selangor, Malaysia

A. A. Jemain
e-mail: azizj@ukm.my

K. Ibrahim
e-mail: kamarulz@ukm.my

1 Introduction

With the development of industrialization and the rapid growth of population, the management of water resources is becoming more important not only in Malaysia but throughout the world. The analysis of rainfall behavior, particularly in terms of the amount of rainfall occurrence is beneficial for managing the consumption of water. Moreover, this analysis could provide information to the experts who will know the spatial and temporal structure of rainfall occurrence in order to give a reliable prediction about the extreme weather events.

The study on the sequence of daily rainfall occurrence was started by Gabriel and Neumann (1962). They found that the daily rainfall occurrence for the Tel Aviv data was successfully fitted with the first-order Markov chain model. Meanwhile, Kottegoda et al. (2004) reported that the first order of the Markov chain model was found to fit the observed data in Italy successfully. This model was based on the assumption that there was a dependency of the daily rainfall occurrence to that of the previous day. Stern and Coe (1984) stated that the two most attractive features of the Markov chain models involved providing the ease in identifying the seasonality in daily rainfall occurrence and

allowing for the flexibility in the building of probability models for the purpose of parameter estimation and model selection.

In most cases, the daily rainfall occurrence can be described by the Markov chain of the first-order model; however, there are cases where this model failed to fit the observed data. As an alternative, the use of the Markov chain model of higher order often improved these inadequacies (see Wilks 1999; Hayhoe 2000). The two criteria that are most commonly used in determining the optimum order for the Markov chain models of higher order are Akaike's information criteria (AIC; see Akaike 1974; Tong 1975) and Bayesian information criteria (BIC; see Schwarz 1978; Katz 1981). The identification of the optimum order of the Markov chain models for the daily rainfall data has been discussed in many studies. Chin (1977), for example, has reported that the optimum order of the Markov chain model was found to vary according to seasons and geographical regions, where the first order was found optimum for summer, while a higher order was observed to be more suitable for winter, based on the fitted models for 100 rainfall stations in the United States. Gates and Tong (1976) had re-examined the Tel Aviv data and revealed that the second-order Markov chain model was found to be optimum instead of the first order. Moreover, Jimoh and Webster (1996) had investigated the optimum order of the daily rainfall occurrence at five stations in Nigeria. Based on the AIC and BIC, they found that the optimum order depends on the threshold levels used and the length of period for the recorded data.

Determining the most appropriate order of the Markov chain model for describing the sequences of wet (dry) days is pertinent in explaining the rainfall behavior. According to Chin (1977), any model which is found to describe the daily rainfall occurrence successfully should also be able to represent the distribution of wet (dry) spells well. The studies on identification of the best order of the Markov chain models to describe the distribution of wet and dry spells have been carried out by a number of researchers at various locations, as in the case of Fleming (2007) and Hui et al. (2005) for Canada; Tolika and Maheras (2005) and Anagnostopoulou et al. (2003) for Greece; Harrison and Waylen (2000) for Costa Rica; Wilks (1999) for the USA; Martin-Vide and Gomez (1999) for Spain; Chapman (1997) for the western Pacific; Dahale et al. (1994) for the tropical South-East Asia and Equatorial Pacific; Moon et al. (1994) for South Korea; Singh et al. (1981) for India; Quah and Ooi (1985) for Peninsular Malaysia; Berger and Goosens (1983) for Belgium; and Basu (1971) for Calcutta.

Although a lot of work on the identification of the best order of the Markov chain models has been done on rainfall data from various parts of the world, only a little attention has been given to the Malaysian data. No study has been conducted to determine the optimum order for the Markov

chain models of the higher order for the daily rainfall in Peninsular Malaysia. Since the study of Quah and Ooi (1985), no further investigation has been done to determine the most appropriate order of the Markov chain models in representing the distribution of wet (dry) spells in this country. Hence, this study is conducted and it is structured as follows. After a brief description of the rainfall data in section Data and methods, the explanation on how to determine the optimum order of the Markov chain for *n*-arbitrary sequences is presented in the section Identifying the optimum order of Markov chain models for daily rainfall occurrences. In the section Persistency of wet (dry) events, the Markov models are described to explain the persistency of the wet (dry) events. This is followed by section Fitting wet (dry) spells using various order of Markov chain models where the most appropriate order of the Markov chain models is identified using the observed distribution of wet (dry) spells data. In sections Results and discussion and Conclusions, the results of the analysis and conclusion of this study are provided respectively.

2 Data and methods

Peninsular Malaysia lies entirely in the equatorial zone, situated in the northern latitude between 1° and 6°N and the eastern longitude from 100° to 103°E. The climate of Peninsular Malaysia is influenced by the Southwest monsoon, from May to August, and the Northeast monsoon, from November until February. The Southwest monsoon is a drier period for the whole country, while during the Northeast monsoon, the east coast and northern areas of Peninsular Malaysia receive more heavy rains than the other parts of the country.

The daily rainfall data from 18 selected stations for Peninsular Malaysia, as provided in Table 1, were selected based on the completeness of the data and the longer period of records. All the data used in this study were collected from the database of the Malaysian Meteorological Services (MMS). The number of years of record for these stations ranged from 21 to 35 years, where 10 stations had records of 35 years, and the rest of the stations had a length of records ranging from 21 to 32 years. The difference in the length of periods for the different stations is due to the availability of the data. Corresponding to the idea by Lim (1976) in this study, the peninsula is divided into five regions, to investigate whether the persistency of wet (dry) days varies according to the different regions of the country. Figure 1 displays a map of Peninsular Malaysia which is divided into the five regions, namely; northwest (NW), west (W), southwest (SW), east (E) and central (C).

In this study, a wet day can be defined as a day when the rainfall amount exceeds or equals a particular threshold

Table 1 Geographic coordinates, altitude and period of records for the selected 18 rain gauged stations in Peninsular Malaysia. *NW* northwest; *SW* southwest; *E* east; *C* central; *W* west

Regions	Rain gauge stations	Latitude	Longitude	Altitude (m)	Period of records
NW	Alor Setar	6°12'N	100°24'E	3.9	1971–2005
	Bayan Lepas	5°18'N	100°16'E	2.8	1971–2005
	Chuping	6°29'N	100°16'E	21.7	1979–2005
W	Ipoh	4°34'N	101°06'E	40.1	1971–2005
	Petaling Jaya	3°06'N	101°39'E	60.8	1971–2005
	Sitiawan	4°13'N	100°42'E	7.0	1971–2005
	Subang	3°07'N	101°33'E	16.5	1971–2005
SW	Kluang	2°01'N	103°19'E	88.1	1974–2005
	Melaka	2°16'N	102°15'E	8.5	1971–2005
	Senai	1°38'N	103°40'E	37.8	1974–2005
E	Kota Bharu	6°10'N	102°17'E	4.6	1971–2005
	Kuala Krai	5°32'N	102°12'E	68.3	1984–2005
	Kuala Terengganu	5°23'N	103°06'E	5.2	1985–2005
	Kuantan	3°47'N	103°13'E	15.3	1971–2005
	Mersing	2°27'N	103°50'E	43.6	1971–2005
C	Batu Embun	3°58'N	102°21'E	59.5	1982–2005
	T.Tinggi Cameron	4°28'N	101°22'E	1,545.0	1983–2005
	Temerloh	3°28'N	102°23'E	39.1	1978–2005

value. A day with the rainfall amount lesser than 0.1 mm is considered as a dry day, when 0.1 mm is assumed as the threshold value. Meanwhile, for the threshold of 10 mm, a dry day can be defined as a day with a small amount of rainfall, less than 10 mm. Otherwise, the day is considered as a wet day. The purpose of applying two different levels of threshold is to investigate whether the optimum order of the Markov chain model is dependent on the level of threshold used.

3 Identifying the optimum order of Markov chain models for daily rainfall occurrences

Let $X_1, X_2, \ldots, X_t \ldots X_n$ denote n binary variables to represent the sequences of wet and dry events in the daily rainfall occurrence for n–arbitrary days, indicated as 1 and 0, respectively. The sequence of a wet (dry) day is assumed to follow a Markov chain model of the first order in the events of wet or dry at time t, when X_t, depends on the previous event X_{t-1}. Let P_{10} denotes the conditional probability of a wet day following a dry day and P_{11} denotes the conditional probability of a wet day following a wet day. Then these two conditional probabilities can be given by

$$P_{10} = P(X_t = 1 | X_{t-1} = 0) \quad \text{and}$$
$$P_{11} = P(X_t = 1 | X_{t-1} = 1) \tag{1}$$

For the Markov chain model of kth order, the stationary transition probabilities are given by

$$P_{i_{k+1},\ldots,i_1} = P(X_t = i_{k+1} | X_{t-1} = i_k, \ldots, X_{t-k} = i_1); \ i = 0, 1 \tag{2}$$

The joint probability distribution for $X_1, X_2, \ldots, X_t, \ldots, X_n$ can be given by

$$P(X_n = i_n, \ldots, X_1 = i_1) = P(i_n, i_{n-1}, \ldots, i_1) \tag{3}$$

Equation (2) can also be written as

$$P(i_n, i_{n-1}, \ldots, i_1) = P(i_n | i_{n-1}, \ldots, i_{n-k}) P(i_{n-1} | i_{n-2}, \ldots, i_{n-k-1}) \ldots$$
$$P(i_{k+1} | i_k, \ldots, i_1) P(i_k | i_{k-1}, \ldots, i_1) \ldots P(i_2 | i_1) P(i_1)$$
$$= P_{i_1} P_{i_2 i_1} \ldots P_{i_k, \ldots, i_1} \prod_{j=1}^{n-k} P_{i_{k+j}, \ldots, i_j} \tag{4}$$

where $P_{i_1} = P(i_1)$.

In order to obtain the optimum order of the Markov chain model, AIC and BIC can be applied. Both criteria are based on the log-likelihood functions for the transition probabilities of the fitted Markov chain models. The maximum likelihood function for the kth order chain is given by

$$L_k(X_1, \ldots, X_n) = \prod_{s_k, \ldots, s_1} \widehat{P}_{s_k, \ldots, s_1}^{n_{s_k, \ldots, s_1}} \tag{5}$$

where, $\widehat{P}_{s_k, \ldots, s_1}^{n_{s_k, \ldots, s_1}}$ denotes the estimated transition probabilities of each of the sequence going from state s_1 to s_2, from state s_2 to s_3, and out to state s_{k-1} to s_k, where s_k is the state of the most recent observation; and the superscript $n_{s_k, \ldots, s_1}$
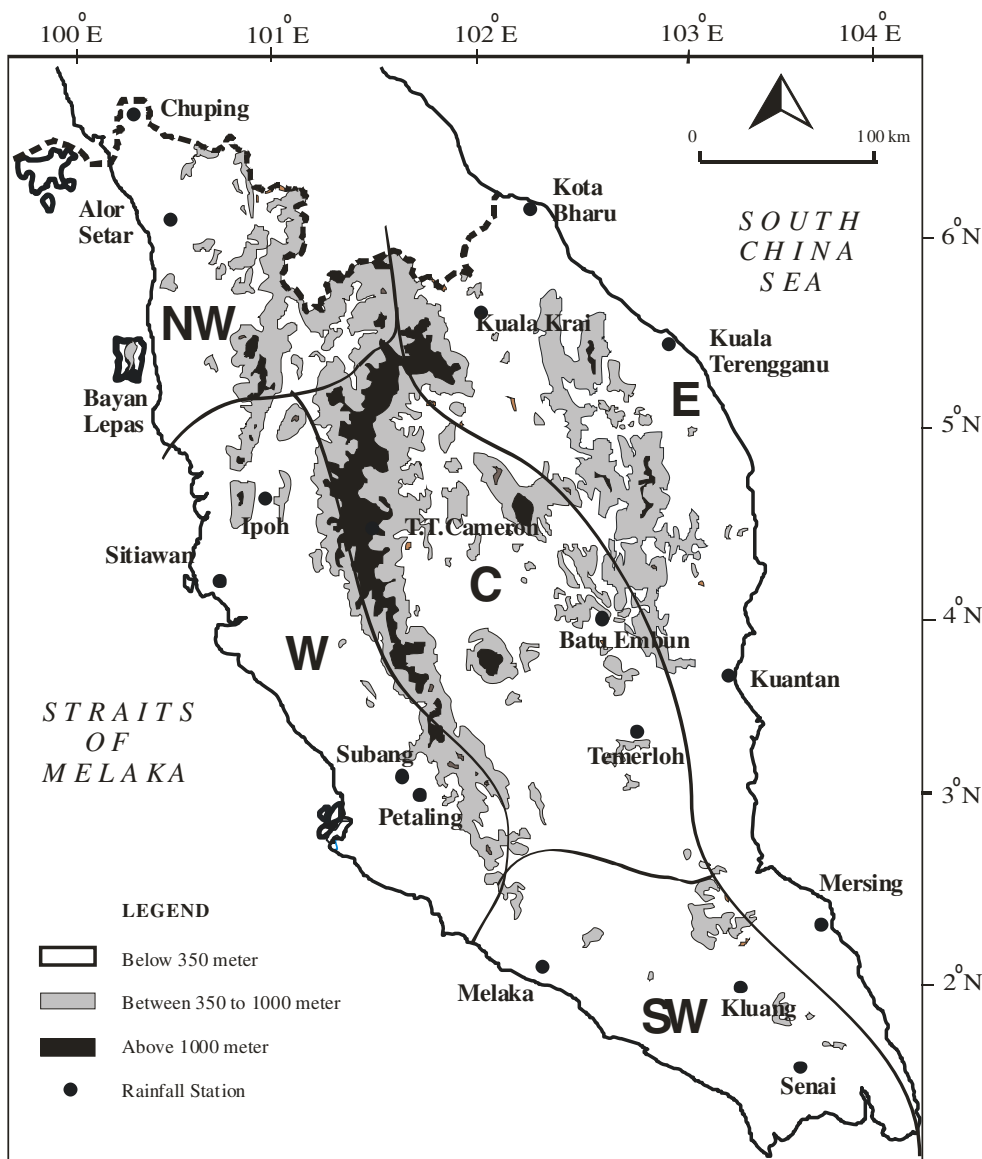
denotes the associated transition counts. The maximum
likelihood estimator used in Eq. (5) of the transition
probabilities of Eq. (2) is given by,

$$\widehat{P}_{s_k,...,s_1} = \frac{n_{s_k,...,s_1}}{n_{\bullet s_{k-1},...,s_1}} \qquad (6)$$

with $n_{\bullet s_{k-1},...,s_1} = \sum_{s_k} n_{s_k,...,s_1}$

The comparison of the two different Markov chain models
to decide on the optimum order, say the Markov chain models
of kth and r^{th} order where $k<r$ and $k=0,1,...,r-1$ can be based
on the maximized likelihood ratio statistics which are
computed as

$$\eta_{k,r} = -2\ln\lambda_{k,r} \qquad (7)$$

where $\lambda_{k,r} = \frac{L_k(X_1,...,X_n)}{L_r(X_1,...,X_n)}$

The selection of the Markov chain model of the optimum
order, say k^{th} order, involves the use of minimum loss

function as obtained from the AIC or BIC. When assessing
the kth order of the Markov chain model, Tong (1975)
proposed the loss function as denoted by AIC (k) to define
the risk on the basis of the AIC criteria which can be given
by

$$\text{AIC}(k) = \eta_{k,r} - 2v \qquad (8)$$

where $v = (S^r - S^k)(S-1)$ is the degree of freedom and S
represents the number of states. Schwarz (1978) introduced a
different form of loss function to define the risk on the basis
of the BIC criteria, denoted as BIC(k), which can be given by

$$\text{BIC}(k) = \eta_{k,r} - v\ln(n) \qquad (9)$$

where n is the sample size. The only difference between the
two criteria is the form of the penalty function. These criteria
attempt to find the value of k that minimizes the loss func-
tions. When identifying the optimum order of the Markov

chain models in the daily rainfall occurrences, these criteria were also applied by many researchers (e.g., Chin 1977; Katz 1981; Wilks 1995; Jimoh and Webster 1996; Harrison and Waylen 2000; Kottegoda et al. 2004; Hui et al. 2005).

4 Persistency of wet (dry) events

The existence of persistence in the sequences of weather events can be measured by Besson's coefficient of persistence (BCP) which can be defined as follows:

$$\text{BCP} = \frac{1 - P_1}{1 - P_{11}} - 1 \tag{10}$$

where P_1 is the probability of a wet day and P_{11} is the conditional probability of a wet day given that the previous day was also a wet day. The positive value of BCP indicates the occurrence of a wet (or dry) event followed by an immediate preceding event. A zero BCP value reflects that there is no persistence for a similar event. The details of the derivation of BCP was explained by Besson (1924) based on an analysis of the persistence of the daily rainfall sequences in Paris. This method was applied by many other researchers (e.g., Erikson 1965; Singh et al. 1981; Berger and Goosens 1983; and Dahale et al. 1994).

5 Fitting wet (dry) spells using various order of Markov chain models

A wet (dry) spell is defined as a period of consecutive days of exactly, say, n wet (dry) days, occurring exactly before a dry (wet) day and returning to the wet (dry) condition in the $(n + 1)^{\text{th}}$ day. The first-order Markov chain only takes into account the condition of the state, either wet or dry, for one preceding day. Similarly, the second order considers the states of the two preceding days and so on. The Markov chain models up to order four has been applied in this present study. The joint probabilities of the kth order of the Markov chain models, are defined as,

$$P(i_n, \ldots, i_2 | i_1) = P_{i_2 i_1} \ldots P_{i_{k-1} \ldots i_1} \prod_{j=1}^{n-k} P_{i_{k+j}, \ldots, i_j}, \tag{11}$$

$$n \geq k + 1; \quad k = 0, 1 \ldots$$

The conditional probabilities of a spell of n wet days under the kth order Markov chain, which is denoted as $P(0, [n]|0)$, are given by,

$$P(0, [n]|0) = \begin{cases} \prod_{i=1}^{n} P_{[i]0} P_{0[n]0} & , \quad n \leq k - 1 \\ \prod_{i=1}^{k} P_{[i]0} P_{[k+1]}^{n-k} P_{0[k]} , & n \geq k \end{cases} \tag{12}$$

where $[n]$ represents the n times. For example, the conditional probability of two consecutive wet days can be written as $P(011|0)$.

In order to compute the expected number of wet days, the conditional probabilities according to the respective length, say, 1, 2, 3, …, n days, which is obtained from Eq. (12) are multiplied by the total number of dry days. The Kolmogorov-Smirnov goodness-of-fit test is employed to compare the observed distribution and the expected distributions of the wet spells for the first few orders of the Markov chain models. The maximum absolute difference between the two cumulative values of the observed and expected number of wet days under the assumed Markov chain models are computed and if these values are found to be less than or equal to the critical value $D_{0.05}$, the particular Markov chain models are considered as most appropriate for describing the distribution of the wet spells. A similar approach is applied to the sequences of dry days.

6 Results and discussion

6.1 Characteristics of the sequence of wet (dry) days

The conditional probabilities of a wet (dry) event at each station are estimated for both levels of threshold and averaged out according to the regions and monsoon seasons. In Table 2, it is found that for both the monsoon seasons and levels of threshold, the conditional probability of a wet day given the previous day was wet, P_{11}, is substantially higher than the probability of wet, P_1, for all the stations. When the levels of persistence of wet spells are compared based on the two levels of threshold, the persistence level is found to be higher for the threshold level of 0.1 mm. For threshold 10 mm, the difference between the successive conditional probabilities of wet days reduces dramatically and becomes negligible after about 2 or 3 days; however, this difference reduces after about 4 or 5 days for the threshold 0.1 mm. Similarly, the conditional probabilities of dry events show that the probability P_{00} is substantially higher than P_0. The difference between the successive conditional probabilities of dry days reduces progressively and becomes negligible after about 5 days and about 2 or 3 days for thresholds 0.1 and 10.0 mm respectively. As expected, the lower level of threshold value will produce a slightly longer duration of wet (dry) periods. The analysis of intense wet spells (e.g., threshold 10 mm) shows that the conditional probabilities of the wet spells, P_{11}, P_{111}, P_{1111} and P_{11111} are lower than those of the dry spells, P_{00}, P_{000}, P_{0000} and P_{00000} during both the monsoon seasons at all regions. In addition, the conditional probabilities of P_{1111} and P_{11111} are found decreased over the southwestern, eastern and the central parts of Peninsular Malaysia for the Southwest monsoon and also in the

Table 2 Persistency characteristics of daily rainfall occurrence for both monsoons at five regions in Peninsular Malaysia during the period of records for both levels of threshold, 0.1 mm[a] 10.0 mm[b]. *NW* northwest; *SW* southwest; *E* east; *C* central; *W* west

	Northeast monsoon					Southwest monsoon					Annual				
	NW	W	SW	E	C	NW	W	SW	E	C	NW	W	SW	E	C
P_1[a]	0.33	0.55	0.48	0.51	0.55	0.54	0.46	0.51	0.46	0.50	0.46	0.52	0.51	0.49	0.54
P_{11}	0.56	0.65	0.64	0.72	0.70	0.65	0.57	0.58	0.53	0.58	0.64	0.63	0.61	0.63	0.65
P_{111}	0.59	0.68	0.66	0.75	0.72	0.66	0.60	0.59	0.54	0.60	0.66	0.66	0.63	0.67	0.68
P_{1111}	0.60	0.70	0.67	0.77	0.74	0.66	0.63	0.59	0.55	0.61	0.66	0.69	0.64	0.70	0.70
P_{11111}	0.61	0.72	0.69	0.79	0.75	0.66	0.66	0.60	0.56	0.63	0.67	0.71	0.66	0.73	0.71
P_0	0.67	0.45	0.52	0.49	0.45	0.46	0.55	0.49	0.54	0.50	0.54	0.48	0.49	0.51	0.46
P_{00}	0.78	0.58	0.68	0.71	0.64	0.59	0.65	0.56	0.60	0.60	0.68	0.60	0.60	0.65	0.60
P_{000}	0.83	0.63	0.73	0.77	0.69	0.62	0.69	0.59	0.63	0.66	0.74	0.65	0.66	0.70	0.66
P_{0000}	0.85	0.66	0.76	0.80	0.72	0.64	0.71	0.61	0.65	0.70	0.78	0.68	0.69	0.74	0.70
P_{00000}	0.87	0.67	0.77	0.82	0.74	0.66	0.72	0.64	0.68	0.73	0.81	0.70	0.72	0.77	0.72
BCP	0.54	0.29	0.46	0.74	0.49	0.31	0.28	0.16	0.15	0.24	0.47	0.29	0.28	0.39	0.34
Average wet spell	2.28	2.85	2.81	3.55	3.35	2.86	2.35	2.38	2.12	2.63	2.75	2.73	2.61	2.70	3.04
Average dry spell	4.83	2.39	3.10	3.53	2.75	2.44	2.83	2.27	2.51	2.52	3.16	2.49	2.52	2.88	2.51
P_1[b]	0.10	0.21	0.16	0.20	0.18	0.20	0.16	0.17	0.16	0.18	0.17	0.19	0.17	0.18	0.19
P_{11}	0.22	0.27	0.30	0.48	0.30	0.30	0.23	0.20	0.20	0.21	0.29	0.27	0.25	0.34	0.27
P_{111}	0.23	0.33	0.31	0.54	0.33	0.33	0.27	0.21	0.20	0.23	0.31	0.32	0.26	0.44	0.31
P_{1111}	0.23	0.35	0.33	0.57	0.33	0.36	0.31	0.19	0.17	0.32	0.34	0.34	0.29	0.51	0.33
P_{11111}	0.11	0.37	0.32	0.58	0.40	0.37	0.33	0.17	0.15	0.13	0.36	0.38	0.29	0.55	0.30
P_0	0.90	0.79	0.84	0.80	0.82	0.80	0.85	0.83	0.84	0.82	0.83	0.81	0.83	0.82	0.81
P_{00}	0.91	0.81	0.87	0.87	0.85	0.82	0.86	0.83	0.85	0.83	0.86	0.83	0.84	0.86	0.83
P_{000}	0.92	0.82	0.88	0.89	0.86	0.83	0.87	0.84	0.85	0.84	0.87	0.83	0.85	0.87	0.84
P_{0000}	0.93	0.82	0.89	0.90	0.87	0.83	0.87	0.84	0.85	0.84	0.88	0.84	0.86	0.87	0.85
P_{00000}	0.93	0.83	0.89	0.91	0.87	0.84	0.88	0.85	0.86	0.85	0.88	0.85	0.86	0.88	0.85
BCP	0.15	0.09	0.20	0.54	0.17	0.15	0.10	0.04	0.05	0.04	0.17	0.11	0.10	0.25	0.11
Average wet spell	1.28	1.38	1.43	1.92	1.44	1.43	1.31	1.26	1.25	1.28	1.40	1.37	1.33	1.53	1.39
Average dry spell	11.68	5.46	7.64	7.92	6.51	5.79	7.21	6.13	6.57	6.08	7.09	5.88	6.44	7.05	5.99

[a] 0.1 mm
[b] 10.0 mm

northwestern areas during the Northeast monsoon. This may be due to the fact that the intense wet spells seldom lasts for more than four consecutive days in most regions over the peninsula. Thus, it is evident to consider the fourth order Markov ($r=4$) as the maximum order of the Markov process. According to Chin (1977), the occurrence of wet (dry) events is highly unlikely to have more than the fourth-order conditional dependence.

Further analysis of the persistency of the wet (dry) events will be explored by applying Besson's coefficient of persistence (BCP). When the Northeast monsoon is considered, the results in Table 2 indicate that the BCP value for the stations located in the eastern region is the highest (0.74) when compared to the other regions; however, it shows the lowest value of 0.15 during the Southwest monsoon for both levels of threshold. For a lower level of threshold used, the average length of wet and dry spells during both monsoon seasons at all regions ranges between 3 to 4 days and 3 to 5 days, respectively. However, for a higher level of threshold used, the average length of dry spells of 6–12 days, is observed to be much longer than that of the wet spells, which is 1–2 days.

Obviously, the results indicate that the persistence of dry sequences during the Northeast monsoon at the northwestern areas is much longer if compared to the other regions. The results of the study seem to be in agreement with the findings of Dale (1960), Camerlengo (1999) and Deni et al. (2008) who found that the dry spells were largely dependent on latitude and were longer and more frequent in the north than in the south and were more frequent during the Northeast monsoon.

6.2 The optimum order of Markov chain models

In determining the optimum order of the Markov chain models for a 5-day sequence in the daily rainfall occurrence, the minimum loss function obtained from the two decision criteria, namely; the AIC and BIC, will be applied. As an example, Table 3 illustrates that the loss function obtained from the AIC and BIC (see Eqs. 8 and 9) consistently shows that the third order is found optimum in almost all the stations for threshold 0.1 mm. However, for a higher level of threshold used, e.g. 10.0 mm, based on the BIC, the

Table 3 The loss function obtained from AIC[a] and BIC[b] for the Markov chain model of kth order, with $r=4$

Regions	Rain guage stations	Threshold = 0.1 mm				Threshold = 10.0 mm			
		0	1	2	3	0	1	2	3
NW[a]	Alor Setar	2085.44	605.35	288.94	115.31[c]	369.05	93.33	42.23	18.12[c]
	Bayan Lepas	1345.65	312.09	127.17	61.63[c]	249.38	56.17	29.01	11.31[c]
	Chuping	1525.14	402.66	177.22	64.89[c]	230.19	52.08	17.93	6.03[c]
W	Ipoh	904.66	239.63	95.66	46.17[c]	155.59	79.85	34.58	26.64[c]
	Petaling	835.94	204.51	82.62	32.43[c]	218.66	70.51	40.04	15.69[c]
	Sitiawan	988.38	261.67	122.43	57.89[c]	116.16	36.47	8.39	2.03[c]
	Subang	904.27	255.69	120.77	53.07[c]	164.67	57.59	18.62	10.46[c]
SW	Kluang	745.10	221.67	75.38	31.71[c]	174.78	48.20	22.71	3.98[c]
	Melaka	645.71	124.77	30.65	16.39[c]	123.59	47.09	37.54	2.12[c]
	Senai	781.40	199.78	66.15	14.66[c]	108.65	21.25	5.79	3.27[c]
E	Kota Bharu	1585.25	335.31	128.12	40.84[c]	855.60	209.90	78.57	32.70[c]
	Kuala Krai	828.33	238.10	108.43	37.42[c]	339.57	84.21	23.54	18.88[c]
	K. Terenggano	1007.20	296.40	130.61	58.74[c]	470.07	122.52	57.84	5.26
	Kuantan	1141.34	285.64	109.29	63.28[c]	442.65	93.80	41.41	8.40[c]
	Mersing	1127.87	359.82	160.62	53.67[c]	414.42	133.25	37.89	3.78[c]
C	Batu Embun	577.51	190.63	77.43	28.42[c]	54.60	9.59	−11.59[c]	−9.15
	T.Tinggi Cameron	1153.84	201.57	67.83	15.08[c]	288.32	88.68	39.95	4.97[c]
	Temerloh	560.29	160.69	45.25	8.06[c]	54.57	22.68	−6.22[c]	−3.86
NW[b]	Alor Setar	1963.20	491.26	191.15	50.11[c]	246.81	−20.76	−55.56[c]	−47.08
	Bayan Lepas	1223.41	198.00	29.38	−3.57[c]	127.14	−57.92	−68.79[c]	−53.88
	Chuping	1406.79	292.20	82.54	1.77[c]	111.84	−58.38	−76.75[c]	−57.09
W	Ipoh	782.42	125.54	−2.13	−19.02[c]	33.35	−34.24	−63.21[c]	−38.55
	Petaling	713.70	90.42	−15.17	−32.77[c]	96.42	−43.58	−57.75[c]	−49.50
	Sitiawan	866.14	147.58	24.63	−7.30[c]	−6.08	−77.63	−89.40[c]	−63.16
	Subang	782.03	141.60	22.98	−12.12[c]	42.43	−56.50	−79.18[c]	−54.74
SW	Kluang	624.21	108.83	−21.34	−32.77[c]	53.88	−64.63	−74.00[c]	−60.49
	Melaka	573.47	10.68	−67.15	−48.81	1.34	−67.00[c]	−60.26	−63.08
	Senai	660.54	86.98	−30.54	−49.80[c]	12.21	−91.55[c]	−90.89	−61.19
E	Kota Bharu	1463.01	221.22	30.33	−24.35[c]	733.36	95.81	−19.22	−32.50[c]
	Kuala Krai	713.69	131.10	16.72	−23.72	224.93	−22.79	−68.17[c]	−42.26
	K. Terenggano	892.62	189.45	38.95	−2.37[c]	355.49	15.58	−33.82	−55.85[c]
	Kuantan	1019.10	171.55	11.50	−1.91[c]	320.40	−20.29	−56.38	−56.79[c]
	Mersing	1005.63	245.73	62.83	−11.52[c]	292.17	19.16	−59.90	−61.41[c]
C	Batu Embun	461.16	82.05	−15.64	−33.63[c]	−61.75	−99.00	−104.67[c]	−71.20
	T.Tinggi Cameron	1038.06	93.50	−24.79	−46.68[c]	172.54	−19.38	−52.68	−56.78[c]
	Temerloh	441.71	50.01	−49.61	−55.18[c]	64.01	−87.99	−101.09[c]	−67.10

[a] AIC

[b] BIC

[c] Indicates the minimum value which corresponds to the optimum order of the 5-day sequence in daily rainfall occurrence over Peninsular Malaysia

optimum order for the stations located in the northwestern and western areas is found to be slightly lower than that of the AIC. For the rest of the analysis, Table 4 displays the summary of the optimum order for a 5-day sequence of wet days at each region considering both the levels of threshold used and both the monsoon seasons. It can be seen that the optimum order obtained from both, decision criteria varies with season, region and threshold used. Generally, for both the AIC and BIC and the thresholds considered, the Markov chain of higher order is found to be optimum at most stations during the Northeast monsoon. Moreover, regardless of the

monsoon seasons, for the two levels of threshold used, the results show that the Markov chain model order higher than one is found to be optimum at all stations in the peninsula. However, for the Southwest monsoon, the results show that the zero and the first-order model are found to be optimum at most of the stations located in the northwestern and eastern areas, when the higher level of threshold of 10 mm, is considered. It is also remarkable that the optimum order estimated by the AIC is either greater or equals to the BIC at most stations for thresholds 0.1 and 10 mm. The findings of this present study is supported by Katz (1981) who claimed

Table 4 The optimum order of Markov chain models for the 5-day sequence at five regions in Peninsular Malaysia over the period of records. *NW* northwest; *SW* southwest; *E* east; *C* central

Region	Rain guage stations	Akaike's information criteria (AIC)						Bayesian information criteria (BIC)					
		Threshold = 0.1 mm			Threshold = 10.0 mm			Threshold = 0.1 mm			Threshold = 10.0 mm		
		Northeast	Southeast	Annual	Northeast	Southeast	Annual	Northeast	Southeast	Annual	Northeast	Southeast	Annual
NW	Alor Setar	>2	>2	>2	>2	1	>2	>2	2	>2	>2	1	2
	Bayan Lepas	>2	2	>2	>2	1	>2	>2	2	>2	1	1	2
	Chuping	>2	2	>2	>2	1	>2	>2	1	>2	>2	1	2
W	Ipoh	2	>2	>2	2	2	>2	2	2	>2	2	2	2
	Petaling	>2	>2	>2	2	>2	>2	2	2	>2	1	2	2
	Sitiawan	>2	>2	>2	>2	1	>2	>2	2	>2	2	1	2
	Subang	>2	>2	>2	>2	2	>2	2	2	>2	2	2	2
SW	Kluang	>2	>2	>2	>2	>2	>2	2	2	>2	2	2	2
	Melaka	>2	2	>2	>2	>2	>2	2	1	2	2	1	>2
	Senai	>2	>2	>2	>2	>2	>2	2	2	>2	2	2	2
E	Kota Bharu	>2	2	>2	>2	>2	>2	>2	2	>2	>2	1	>2
	Kuala Krai	>2	>2	>2	2	0	>2	2	1	>2	2	0	2
	K. Terengganu	>2	2	>2	>2	1	>2	2	2	>2	>2	0	>2
	Kuantan	>2	2	>2	>2	0	>2	>2	1	>2	>2	0	>2
	Mersing	>2	>2	>2	>2	2	>2	>2	2	>2	>2	1	>2
C	Batu Embun	>2	>2	>2	2	2	>2	2	2	>2	2	2	2
	T.Tinggi Cameron	>2	>2	>2	>2	>2	>2	2	2	>2	2	1	>2
	Temerloh	>2	>2	>2	>2	0	2	2	2	>2	1	0	2

that the optimum order obtained based on the AIC was inconsistent and had a tendency to over-estimate the true optimum order, whereas the BIC produced more consistent results.

6.3 The most appropriate order for the wet (dry) spells

The spatial distribution of the most appropriate order of Markov chain models to explain the wet (dry) spells (see section Fitting wet (dry) spells using various order of Markov chain models) over Peninsular Malaysia is identified through the Kolmogorov-Smirnov goodness-of-fit test. Figures 2 and 3 provide the observed and expected frequencies for both levels of threshold based on the most appropriate order of the first four orders of the Markov chain models identified as MC1, MC2, MC3 and MC4 for the

distribution of the dry and wet spells for each of the four selected stations, respectively. It is obvious that the expected frequencies of dry and wet spells which are obtained from the most appropriate order of the Markov chain models really describe the observed distributions. It is also found that a higher order of the Markov chain models are observed to be more appropriate in order to describe the distribution of dry spells at the four selected stations, for both levels of threshold considered, as shown in Fig. 2. Meanwhile, Fig. 3 indicates that the first and second order are found to be the most appropriate models to describe the distribution of wet spells at each of the four selected stations.

Figures 4 and 5 display the results of the analyses on the most appropriate order for distribution of the wet and dry spells respectively at each of the selected rainfall stations in

Fig. 2 The observed (*white strip*) and estimated (*black dot*) frequencies of dry spells for thresholds **a** 0.1 and **b** 10 mm based on the most appropriate order of Markov chain models (*MC2*, *MC3* and *MC4*) at each of the four selected rainfall stations
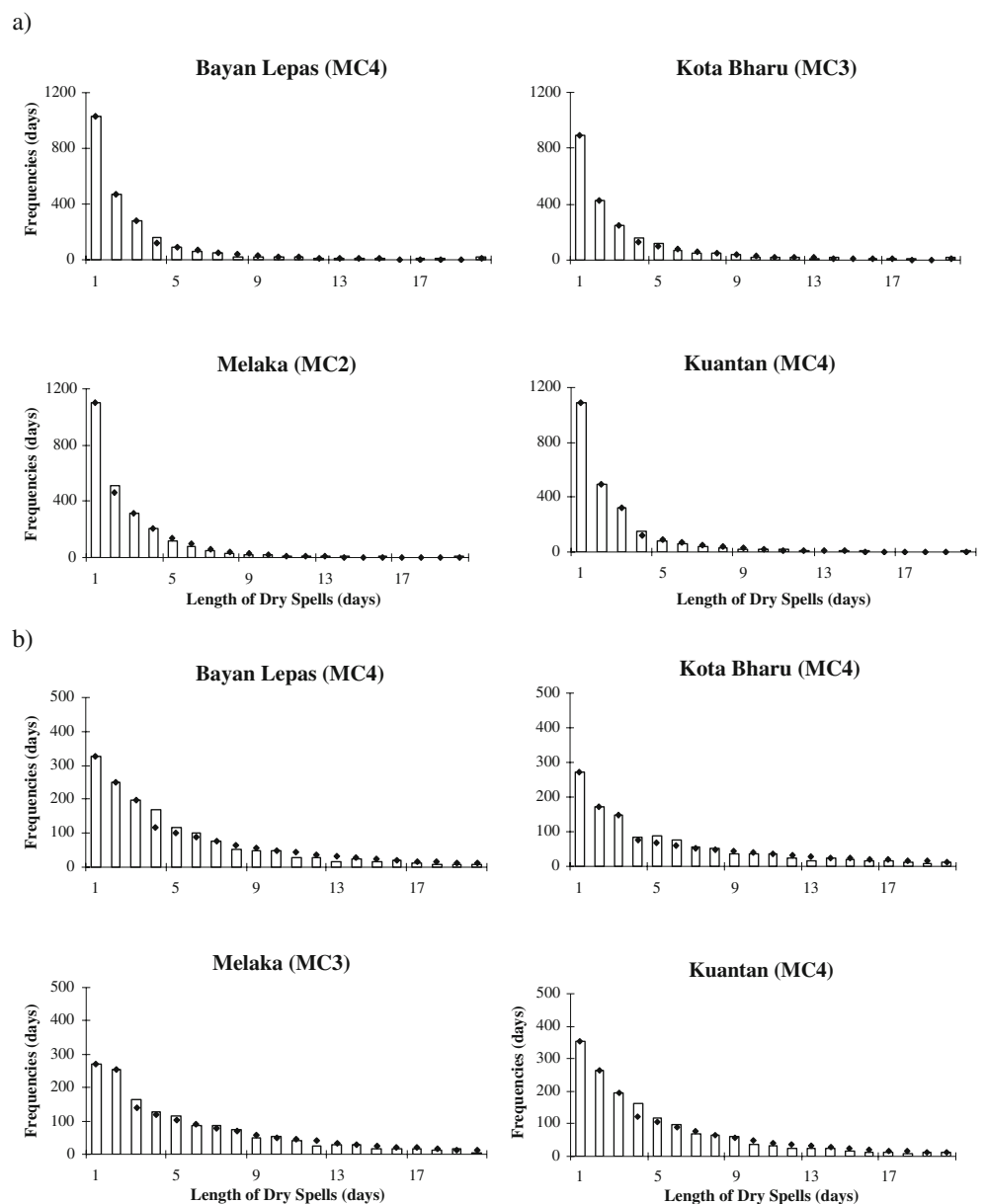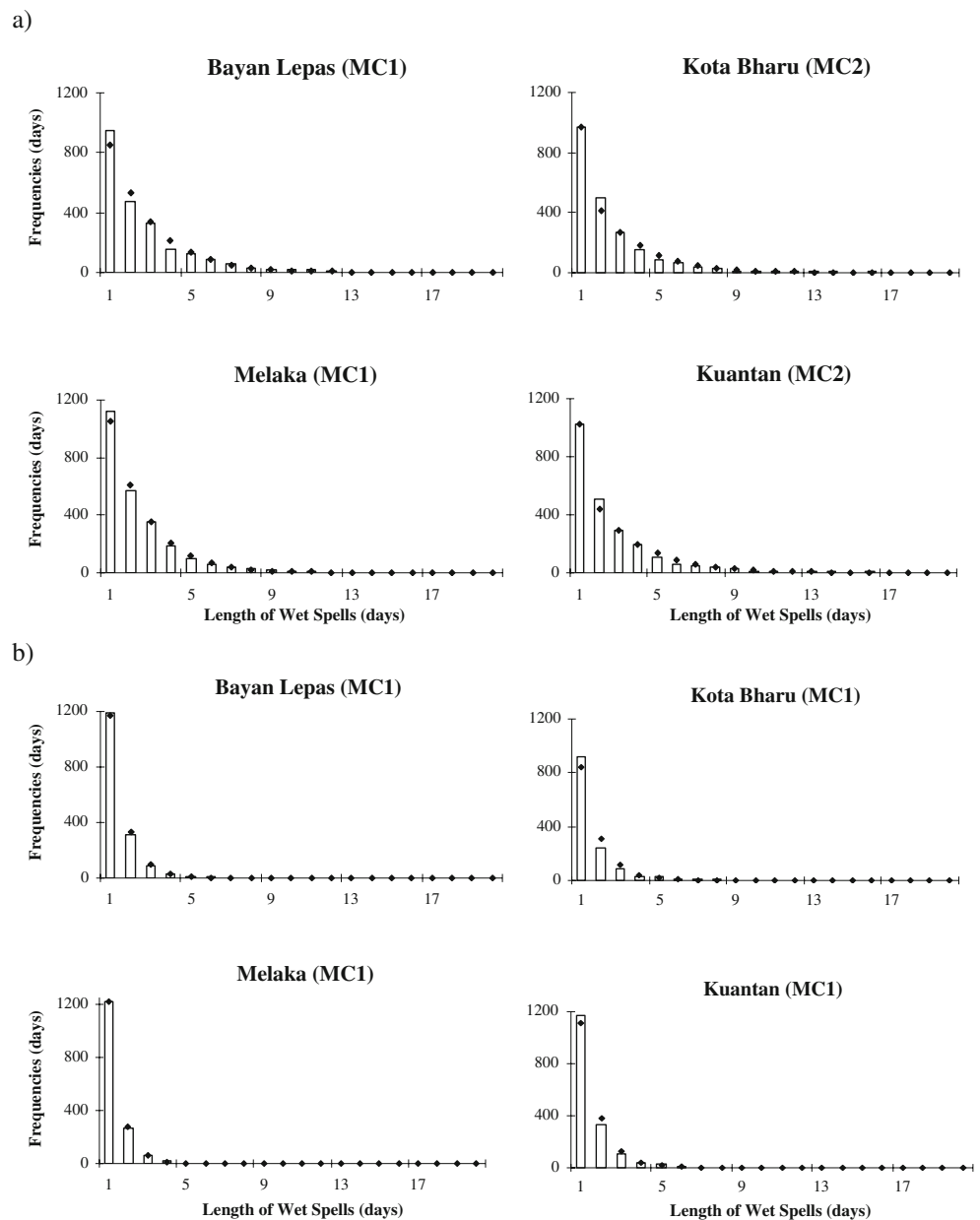
Fig. 3 The observed (*white strip*) and estimated (*black dot*) frequencies of wet spells for thresholds **a** 0.1 and **b** 10 mm based on the most appropriate order of Markov chain models (*MC1* and *MC2*) at each of the four selected rainfall stations

Peninsular Malaysia. As shown in Fig. 4, it is generally found that the Markov chain of higher order is appropriate for describing the distribution of dry spells during the Northeast monsoon at most stations except for the stations at Subang and Petaling. However, during the Southwest monsoon, for both levels of threshold, the first-order model is found appropriate at all stations except for the four stations located in Subang, Sitiawan, Ipoh and Temerloh. Regardless of the monsoon season, the results for both levels of threshold consistently show that the Markov chain of higher order is found to be appropriate for representing the distribution of dry spells at all stations in Peninsular Malaysia.

Figure 5 displays the most appropriate order of the first four orders of the Markov chain models in representing the distribution of wet spells for each of the rainfall stations considered in this study. For the higher level of threshold applied, Fig. 5d–f indicates that the first-order model is found to be the most appropriate in representing the distribution of intense wet spells in most of the stations over the peninsula. Moreover, for the lower threshold level of 0.1 mm, the Markov chain of higher order is found to be more appropriate in the eastern, western and central parts of the peninsula, regardless of the monsoon seasons. Generally, the first-order Markov chain model is found more appropriate in describing the distribution of the wet spells at most regions during the Northeast and Southwest monsoons except in the western areas for the lower level of threshold of 0.1 mm. It can be concluded here that the Markov chain
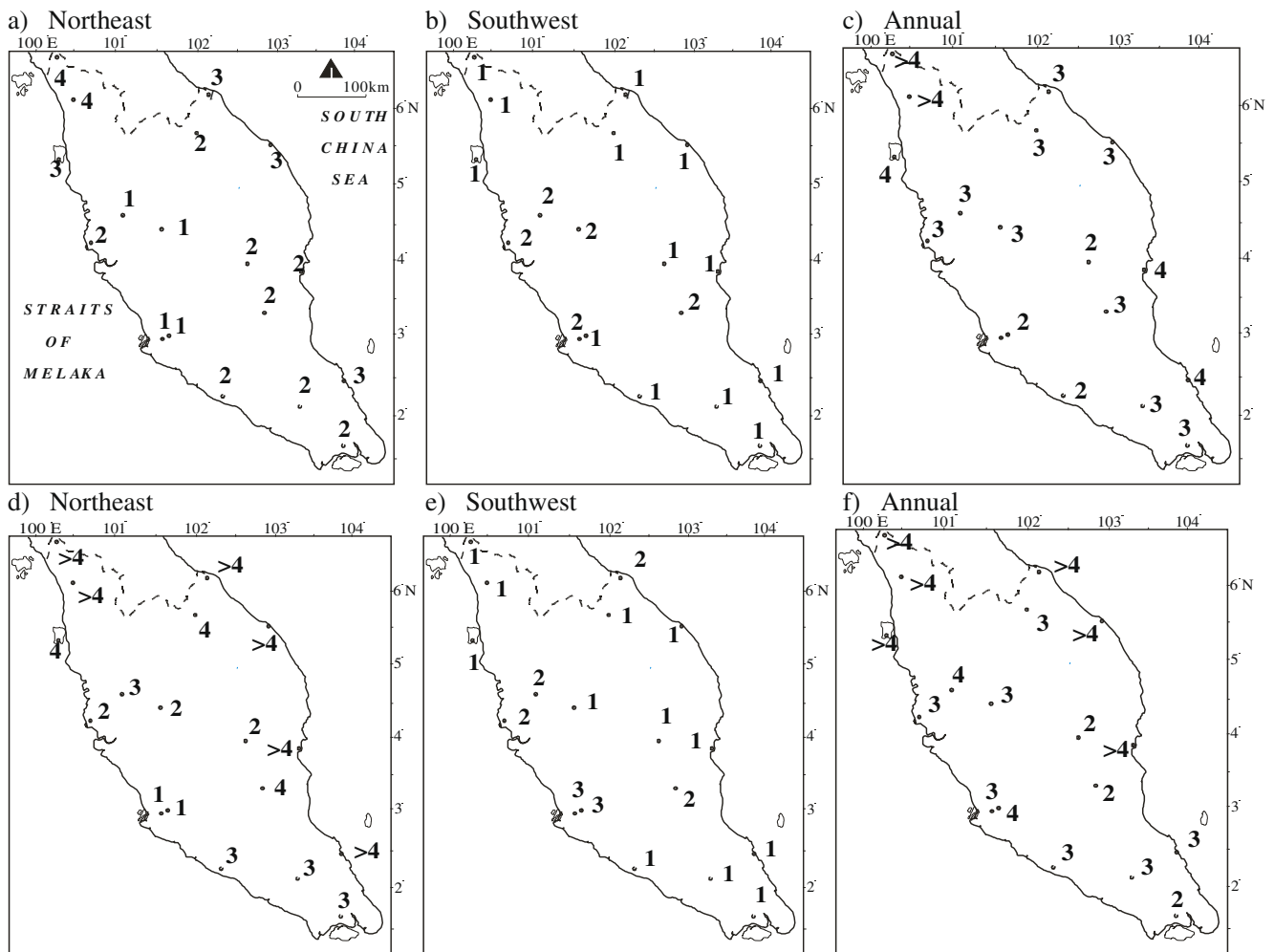
Fig. 4 The most appropriate order of Markov chain models for the distribution of dry spells at all stations in Peninsular Malaysia over the period of records, **a**–**c** for threshold = 0.1 mm and **d**–**f** for threshold 10.0 mm

of higher order is more appropriate for the dry spells than for the wet spells. Similar findings had also been found by Dahale et al. (1994).

7 Conclusions

The information of wet (dry) behavior is vitally important to the water resource management which can be used for the prediction of extreme weather events. The analysis of persistency of wet (dry) spells contributes to the information of the occurrence of a wet or dry event based on the previous day. The results obtained from the analysis will provide a more complete description on the rainfall occurrences behavior in Peninsular Malaysia. In this present study, the Markov chain models of various orders were applied to the daily rainfall data in order to identify the optimum order of the model which was able to represent the wet (dry) spells by

considering the various geographical regions, different levels of threshold used and also the monsoon seasons.

It can be concluded from this study that the optimum order of the Markov chain models for the 5-day of daily rainfall occurrence for Peninsular Malaysia varies according to the levels of threshold used, the regions and types of monsoon seasons. Generally, for all cases, the results show that the optimum order obtained by using the AIC is higher compared to the results of the alternative method BIC. The findings support the results of simulation theory of Katz (1981), which indicated that the AIC produced inconsistent results compared to the BIC. In addition, Jimoh and Webster (1996) also claimed that the estimation of the optimum order using the AIC tended to overestimate the results. Moreover, the order greater than one is more appropriate for Peninsular Malaysia for both thresholds regardless of the monsoon seasons and using both the decision criteria, AIC and BIC. Generally, the lower order of the Markov chain models is found to be
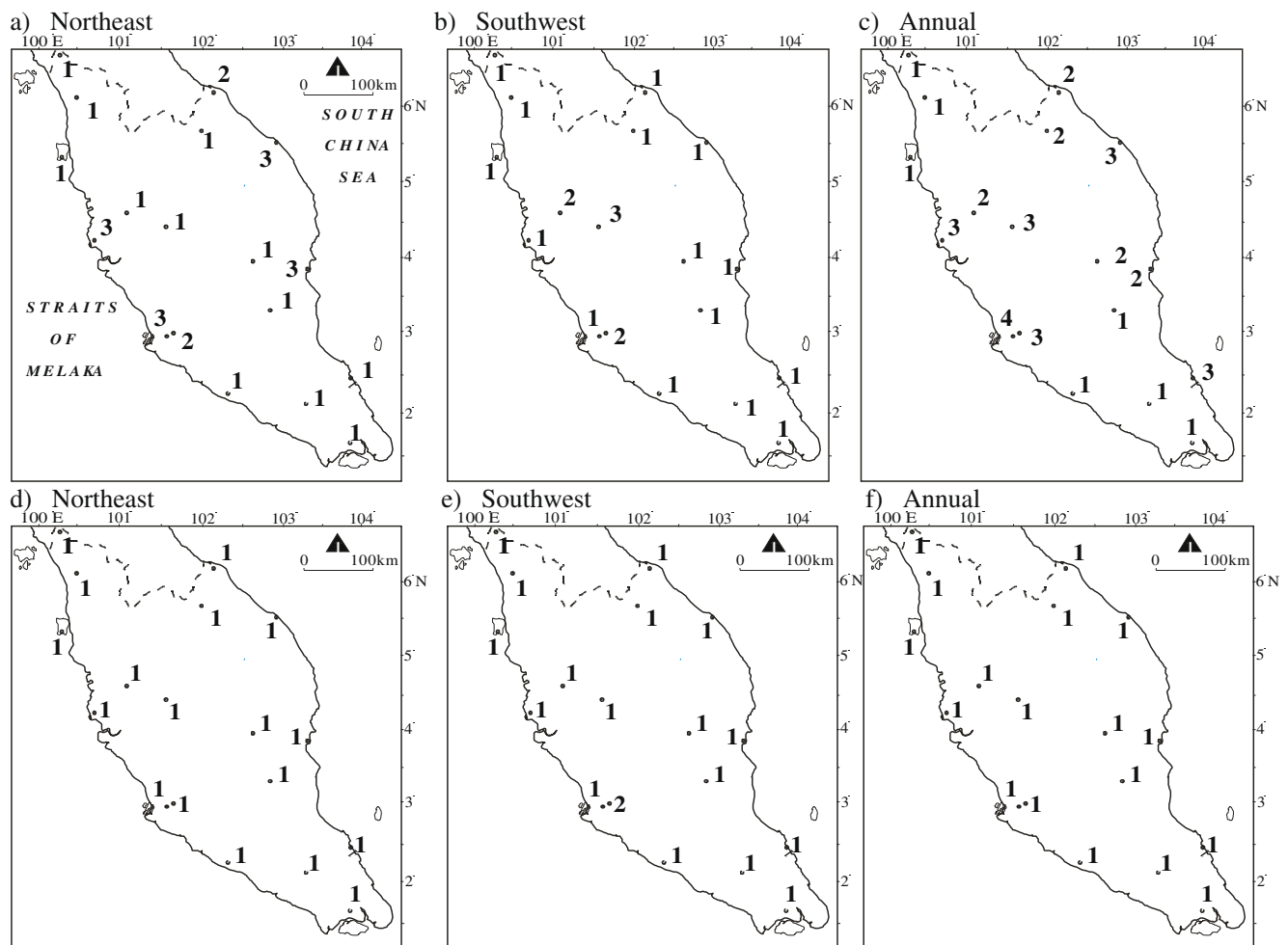
Fig. 5 The most appropriate order of Markov chain models for the distribution of wet spells at all stations in Peninsular Malaysia over the period of records, **a–c**)for threshold = 0.1 mm and **d)f** for threshold 10.0 mm

optimum in the northwestern and the eastern areas during the Southwest monsoon season for a higher level of threshold, 10 mm.

However, the results show that the optimum orders for the Northeast monsoon are equal to or higher than those of the Southwest monsoon, when both levels of threshold are considered. This present study seems to be in agreement with Dahale et al. (1994) who claimed that the closer the stations are to the equator the higher the order of the models. Their findings indicated that the southern stations in Thailand and the equatorial Pacific which are closer to the equator showed a higher order of the Markov chains. Since the higher order of the Markov chain models is found to be optimum for the Northeast monsoon, it is shown in this present study that for the dry spells, the order higher than one is also found to be more appropriate in most of the stations in the peninsula for this monsoon season. This also supports the analysis of persistency which indicates that the BCP of threshold 0.1 mm during the Northeast monsoon is higher when compared to the Southwest monsoon. Thus, the results may provide evi-

dence that the higher the order of the chain, the greater the consideration of the persistence of wet (dry) events. On the other hand, for the wet spells, the first order of the Markov chain model is found to be the most appropriate order in most of the stations over the peninsula for both monsoon seasons and levels of threshold considered.

Further analysis could be carried out by including other rainfall stations in Malaysia so that more valuable information could be found for the purpose of prediction. In the future, the most appropriate probability models by comparing the various orders of the Markov models with the alternative rainfall occurrence models in representing the distribution of wet (dry) spells will be investigated further over Peninsular Malaysia by considering the monsoon seasons and other climatic issues such as global warming or climate change.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723

Anagnostopoulou CHR, Maheras P, Karacostas T, Vafiadis M (2003) Spatial and temporal analysis of dry spells in Greece. Theor Appl Climatol 74(1–2):77–91

Basu AN (1971) Fitting of a Markov chain model for daily rainfall data at Calcutta. Indian J Meteorol Geophys 22:67–74

Berger A, Goosens C (1983) Persistence of wet and dry spells at Uccle (Belgium). J Climatol 3:21–34

Besson L (1924) On the probability of rain. Mon Wea Rev 52:308

Camerlengo AL (1999) Monthly frequency distributions of both dry spells and the number of days with precipitation greater than 25 mm over Peninsular Malaysia. GEOACTA 23:1–18

Chapman TG (1997) Stochastic models for daily rainfall in the Western Pacific. Math Comput Simul 43(3):351–358

Chin EH (1977) Modeling daily precipitation occurrence process in Markov chain. Water Resour Res 13(6):949–956

Dahale SD, Panchawagh N, Singh SV, Ranatunge ER, Brikshavana M (1994) Persistence in Rainfall Occurrence over Tropical South East Asia and Equatorial Pacific. Theor Appl Climatol 49:27–39

Dale WL (1960) The rainfall of Malaya, part II. J Trop Geogr 14:11–28

Deni SM, Jemain AA, Ibrahim K (2008) The spatial distribution of wet and dry spells over Peninsular Malaysia. Theor Appl Climatol (in press). doi:10.1007/s00704-007-0355-8

Eriksson B (1965) A climatological study of persistency and probability of precipitation in Sweden. Tellus 17:484–497

Fleming SW (2007) Climatic influences on Markovian transition matrices for Vancouver daily rainfall occurrence. Atmos Ocean 45(3):163–171

Gabriel KR, Neumann J (1962) A Markov chain model for daily rainfall occurrence at Tel Aviv. Quart J Roy Meteorol Soc 88:90–95

Gates F, Tong H (1976) On Markov chain modeling to some weather data. J Appl Meteorol 15:1145–1151

Harrison M, Waylen P (2000) A note concerning the proper choice for Markov model order for daily precipitation in the humid tropics: a case study in Costa Rica. Int J Climatol 20:1861–1872

Hayhoe HN (2000) Improvements of stochastic weather data generators for diverse climates. Clim Res 14:75–87

Hui W, Xuebin Z, Elaine MB (2005) Stochastic modeling of daily precipitation for Canada. Atmos Ocean 43(1):23–32

Jimoh OD, Webster P (1996) Optimum order of Markov chain for daily rainfall in Nigeria. J Hydrol 185:45–69

Katz RW (1981) On some criteria for estimating the order of Markov chain. Technometrics 23:243–249

Kottegoda NT, Natale L, Raiteri E (2004) Some considerations of periodicity and persistence in daily rainfalls. J Hydrol 296:23–37

Lim JT (1976) Rainfall minimum in Peninsular Malaysia during the Northeast monsoon. Mon Wea Rev 104:96–99

Martin-Vide J, Gomez L (1999) Regionalization of Peninsular Spain based on the length of dry spells. Int J Climatol 19:537–555

Moon SE, Ryoo SB, Kwon JG (1994) A Markov chain model for daily precipitation occurrence in South Korea. Int J Climatol 14(9):1009–1016

Quah LC, Ooi MK (1985) Wet and dry spell probability models in West Malaysia. Research Publication no. 14. Malaysian Meteorological Service, Petaling Jaya, Malaysia

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Singh SV, Kripalani RH, Priya S, Ismail PMM, Dahale SD (1981) Persistence in daily and 5-day summer monsoon rainfall over India. Arch Meteor Geophys Bioclim Ser A 30:261–277

Stern RD, Coe R (1984) A model fitting analysis of daily rainfall data. J R Stats Soc A 147(1):1–34

Tolika K, Maheras P (2005) Spatial and temporal characteristics of wet spells in Greece. Theor Appl Climatol 81(1–2):71–85

Tong H (1975) Determination of the order of a Markov chain by Akaike's information criterion. J Appl Prob 12:488–497

Wilks DS (1995) Statistical methods in atmospheric sciences: an introduction. Academic, London

Wilks DS (1999) Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agric For Meteorol 93(3):153–169